

Detecting non-sinusoidal periodicities in observational data: the von Mises periodogram for variable stars and exoplanetary transits

Roman V. Baluev^{*}

*Central Astronomical Observatory at Pulkovo of Russian Academy of Sciences, Pulkovskoje shosse 65, St Petersburg 196140, Russia
Sobolev Astronomical Institute, St Petersburg State University, Universitetskij prospekt 28, Petrodvorets, St Petersburg 198504, Russia*

Received 2013 February 5; in original form 2012 December 26

ABSTRACT

This paper introduces an extension of the linear least-squares (or Lomb-Scargle) periodogram for the case when the model of the signal to be detected is non-sinusoidal and depends on unknown parameters in a non-linear manner. The attention is paid to the problem of estimating the statistical significance of candidate periodicities found using such non-linear periodograms. This problem is related to the task of quantifying the distributions of maximum values of these periodograms. Based on recent results in the mathematical theory of extreme values of random field (the generalized Rice method), we give a general approach to find handy analytic approximation for these distributions. This approximation has the general form $e^{-z}P(\sqrt{z})$, where P is an algebraic polynomial and z being the periodogram maximum.

The general tools developed in this paper can be used in a wide variety of astronomical applications, for instance in the studies of variable stars and extrasolar planets. For this goal, we develop and consider in details the so-called von Mises periodogram: a specialized non-linear periodogram where the signal is modelled by the von Mises periodic function $\exp(\nu \cos \omega t)$. This simple function with an additional non-linear parameter ν can model lightcurves of many astronomical objects that show periodic photometric variability of different nature. We prove that our approach can be perfectly applied to this non-linear periodogram.

We provide a package of auxiliary C++ programs, attached as the online-only material. They should facilitate the use of the von Mises periodogram in practice.

Key words: methods: data analysis - methods: statistical - surveys

1 INTRODUCTION

The Lomb (1976)-Scargle (1982) (hereafter LS) periodogram is a well-known powerful tool, which is widely used to search for periodicities in observational data. The main idea used in the LS periodogram is to perform a least-squares fit of the data with a sinuous (harmonic) model of the signal and then to check how much the resulting value of χ^2 function improves for a given signal frequency. The maximum value of the LS periodogram (i.e., the maximum decrement in the χ^2 goodness-of-fit measure) corresponds to the most likely frequency of the periodic signal. This natural idea is quite easy to implement in numerical calculations. The linearity of the harmonic model with respect to unknown parameters (two coefficients near the sine and cosine) introduces additional simplifications.

Any signal detecting tool is not of much use without accompanying method of estimating the statistical significance of candidate periodicities. Indeed, the random errors contaminating the input data inspire noise fluctuations on the periodogram, so that we can never be completely sure that the peak that we actually observed is a result of real periodicity in the data. To assess the statistical significance of the observed periodogram peak, we need to calculate the ‘false alarm probability’ (hereafter FAP) associated with this peak. The FAP is the probability that the observed or larger periodogram peak could be produced by random measurement errors. The smaller is FAP, the larger is the statistical significance. Given some small tolerance value FAP_* (say, 1%), we could claim that the detected candidate periodicity is statistically significant (when $FAP < FAP_*$) or is not (when $FAP > FAP_*$).

From the statistical view point, the FAP is tightly connected with the probability distributions of the periodogram

^{*} E-mail: roman@astro.spbu.ru

considered under the null hypothesis (i.e., the hypothesis of no signal in the data). If the frequency of the putative signal was known a priori, we could use only single value of the LS periodogram to check whether the presence of this periodicity is likely or not. In this case, the FAP is given by the well-known exponential distribution of any single value z of the LS periodogram, so that $\text{FAP} = e^{-z}$. However, the case in which the frequency of possible signal is basically unknown is much more common. In this case, the FAP is provided by the distribution function of the maximum value of the periodogram (corresponding to the frequency range being scanned).

The calculation of the latter distribution is a non-trivial task. The absence of accurate and/or rigorous analytic expression of this distribution (even for the plain LS periodogram) represented a significant trouble for astronomers for about three decades. In addition to the Lomb and Scargle works, it is worthwhile to mention here the papers by Horne & Baliunas (1986); Koen (1990); Schwarzenberg-Czerny (1998a); Cumming et al. (1999); Cumming (2004); Frescura et al. (2008). We believe that this obstacle was the main reason why basically no intricate extensions of the LS periodogram attained enough practical popularity so far. Theoretically, it is not really difficult to construct a periodogram where some fancy models of the data are used. Armed with modern computers, we even may evaluate such periodograms in practice, even if they rely on some CPU-greedy numerical algorithms. But what to do next? How to decide which of the signals detected are real and which belong to the noise? The only general solution available is the Monte Carlo simulation technique, which might be practically useful for the basic LS periodogram, but not for more complicated cases, unfortunately.

Rather recently, a significant progress in this field was attained in the paper (Baluev 2008), where entirely analytic and simultaneously accurate approximations of the FAP are given, based on the results in the theory of extreme values of stochastic processes (the ‘Rice method’). In a brief form, the main result presented in (Baluev 2008) for the LS periodogram is:

$$\text{FAP}(z) \lesssim M(z) \approx W e^{-z} \sqrt{z}, \quad (1)$$

where z is the maximum periodogram value corresponding to a given frequency range, and W is the width of this range multiplied by a certain effective length of the time series (which is usually close to the plain time span). The symbol ‘ \lesssim ’ in (1) means that $\text{FAP}(z)$ will never exceed $M(z)$ and simultaneously $M(z)$ represents an asymptotic approximation for $\text{FAP}(z)$, with the error decreasing for small FAP (or large z). The high practical importance of the approximation in (1) is founded on three things: (i) it is entirely analytic, eliminating any need for Monte Carlo simulations, (ii) its practical accuracy usually appears good or at least quite satisfactory, and (iii) its possible errors never favour to more false alarms than we expect, since we deal with an upper limit on FAP.

The LS periodogram can be easily generalized in multiple ways to encompass more complicated models (Schwarzenberg-Czerny 1998a,b; Baluev 2008; Zechmeister & Kürster 2009; Ferraz-Mello 1981). First, we can introduce some base model of an expected underlying variation (typically a long-term polynomial trend) and

check whether the addition of a probe sinusoidal signal offers enough improvement in χ^2 . These cases have been briefly considered in (Baluev 2008) and our general conclusion was that such a modification does not typically break the result (1). Second, we can deal with more complicated (but still linear) models than just a sinusoid. In particular, in the work (Baluev 2009b) we considered the so-called multi-harmonic periodograms, where the periodic signal is modelled by a trigonometric polynomial involving a few leading terms of the Fourier series (Schwarzenberg-Czerny 1996). In this case, the formula (1) is generalized to

$$\text{FAP}(z) \lesssim M(z) \approx W \alpha_n e^{-z} z^{n-1/2}, \quad (2)$$

where α_n are certain numbers depending on the degree n of the approximating trigonometric polynomial. Notice that $n = 1$ corresponds to the LS case.

However, non-sinusoidal periodic signals, which are dealt with in astronomy, often obey non-linear models. Then the use of the LS periodogram or periodograms from (Baluev 2008, 2009b) is not optimal, since the corresponding periodic variation might be fitted by an inadequate model. For instance, this is the case for lightcurves of variable stars and for radial velocity curves of spectral binaries involving elongated orbits. Theoretically, we could use a high-order Fourier expansion to approximate a non-sinusoidal periodicity, but this solution is obviously inefficient due to an unnecessarily large number of extra free parameters. The aim of the present paper is to extend the results from (Baluev 2008) and (Baluev 2009b) to the case of an arbitrary model of the periodic signal, incorporating a few parameters in a non-linear manner. As we will demonstrate, we can apply roughly the same technique (the Rice method) to this case, with the major difference that we should now deal with random fields instead of random processes. Namely, we will provide a closed approach to construct the limiting approximation $M(z)$ in the form $W e^{-z} P(\sqrt{z})$, where P is an algebraic polynomial.

The structure of the paper is as follows. In Section 2, we introduce a general definition extending the LS periodogram to the non-linear case. In Section 3, we consider the problem of assessing the statistical significance of candidate periodicities detected with the non-linear periodogram. This description is followed by an auxiliary Section 4 devoted to the ways of practical evaluation of the theoretical approximations of Section 3. In Section 5, we discuss the consequences implied by various noise models of the data. In Section 6, a couple of concrete practical applications of these results is supplied. In the first (rather tutorial) example, we aim to detect a periodic signal of arbitrary (but a priori fixed) shape, when the unknown parameters are the amplitude and the phase of the signal. In the second example, we consider a more complicated periodogram based on the so-called von Mises model of the signal, essentially $\exp(\nu \cos x)$, which involves an additional non-linear parameter ν .

2 DEFINITION OF THE NON-LINEAR PERIODOGRAM

Let x_i denote the outcomes of N observations made at timings t_i . The errors of these measurements are assumed to

follow Gaussian distributions and to be statistically independent (hence, uncorrelated). The standard deviations of these errors, σ_i , are assumed to be known a priori. We want to test, whether these observations are consistent with some base model of variation, or certain deterministic periodicity is also present.

The data model to be tested for consistency with the data is $\mu_{\mathcal{H}}(t, \boldsymbol{\theta}_{\mathcal{H}})$, where the vector $\boldsymbol{\theta}_{\mathcal{H}}$ incorporates $d_{\mathcal{H}}$ unknown parameters, which should be estimated from the data. We assume that this model is linear with respect to unknown parameters:

$$\mu_{\mathcal{H}}(t, \boldsymbol{\theta}_{\mathcal{H}}) = \boldsymbol{\theta}_{\mathcal{H}} \cdot \boldsymbol{\varphi}_{\mathcal{H}}(t), \quad (3)$$

where the vector of base functions $\boldsymbol{\varphi}_{\mathcal{H}}(t)$ is set a priori. Typically, the model $\mu_{\mathcal{H}}$ incorporates a free constant term, and, possibly, a long-term polynomial trend with free coefficients. The Lomb-Scargle periodogram, by the way, assumes that $\mu_{\mathcal{H}} \equiv 0$, implicitly requesting some preliminary centering of the time series.

The model of the periodic signal is given by $\mu(t, \boldsymbol{\theta}, f)$, where the vector $\boldsymbol{\theta}$ contains d unknown parameters to be estimated from the data together with the frequency f . The united vector $\boldsymbol{\theta}_{\mathcal{K}} = \{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}\}$ parametrizes the compound alternative model of the data¹, which is given by

$$\mu_{\mathcal{K}}(t, \boldsymbol{\theta}_{\mathcal{K}}, f) = \mu_{\mathcal{H}}(t, \boldsymbol{\theta}_{\mathcal{H}}) + \mu(t, \boldsymbol{\theta}, f). \quad (4)$$

We denote the d -dimensional domain, where $\boldsymbol{\theta}$ is supposed to reside, by Θ . The signal is supposed to vanish when $\boldsymbol{\theta}$ belongs to some ‘null domain’ $\Theta_0 \subset \Theta$ and not to vanish when $\boldsymbol{\theta}$ does not belong to Θ_0 . Therefore, we wish to test, whether the data are consistent with the base hypothesis $\mathcal{H} : \boldsymbol{\theta} \in \Theta_0$ (implying that the model $\mu_{\mathcal{H}}(t, \boldsymbol{\theta}_{\mathcal{H}})$ fits the data satisfactory) or this base hypothesis should be rejected in favour of the alternative $\mathcal{K} : \boldsymbol{\theta} \in \Theta \setminus \Theta_0$ (implying the model $\mu_{\mathcal{K}}(t, \boldsymbol{\theta}_{\mathcal{K}}, f)$). The model μ may be non-linear with respect to $\boldsymbol{\theta}$.

The unknowns $\boldsymbol{\theta}_{\mathcal{H}}$, $\boldsymbol{\theta}$, and f can be estimated using the least-squares approach. Under the hypothesis \mathcal{H} , the best-fitting estimation of $\boldsymbol{\theta}_{\mathcal{H}}$ can be obtained in result of minimizing the functions $\chi_{\mathcal{H}}^2(\boldsymbol{\theta}_{\mathcal{H}}) = \langle (x - \mu_{\mathcal{H}})^2 \rangle$. Under the hypothesis \mathcal{K} , the best-fitting estimations of $\boldsymbol{\theta}_{\mathcal{H}}$, $\boldsymbol{\theta}$ and f should correspond to the minimum value of the function $\chi_{\mathcal{K}}^2(\boldsymbol{\theta}_{\mathcal{K}}, f) = \langle (x - \mu_{\mathcal{K}})^2 \rangle$. Since $\mu_{\mathcal{H}}$ is linear, the minimizations by $\boldsymbol{\theta}_{\mathcal{H}}$ can be performed rapidly and precisely using the usual linear least-squares algorithms. The minimization of $\chi_{\mathcal{K}}^2$ over the remaining variables is equivalent to the maximization of the non-linear function

$$\zeta(\boldsymbol{\theta}, f) = \frac{1}{2} \left[\min_{\boldsymbol{\theta}_{\mathcal{H}}} \chi_{\mathcal{H}}^2(\boldsymbol{\theta}_{\mathcal{H}}) - \min_{\boldsymbol{\theta}_{\mathcal{H}}} \chi_{\mathcal{K}}^2(\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}, f) \right], \quad (5)$$

which simultaneously characterises the improvement in the χ^2 fit quality, which is achieved by means of adding to the base model the model of the periodic signal with given values of $\boldsymbol{\theta}$ and f . Note that the maxima of $\zeta(\boldsymbol{\theta}, f)$ do not depend

on the choice of the parametrization. That is, they are invariable with respect to a non-degenerated transformation of the vector $\boldsymbol{\theta}$ (and any non-degenerated linear transformation of $\boldsymbol{\theta}_{\mathcal{H}}$).

We can perform the minimization over the frequency f in a traditional manner by means of looking for the highest peak on the graph of the function

$$z(f) = \max_{\boldsymbol{\theta} \in \Theta} \zeta(\boldsymbol{\theta}, f), \quad (6)$$

which may be called the “non-linear least-squares periodogram”. This definition means that for any fixed frequency f we perform the fit of our model via the remaining d parameters $\boldsymbol{\theta}$. The value of the periodogram characterizes the relevant advance in the χ^2 fit quality. When the model $\mu(t, \boldsymbol{\theta}, f)$ is linear with respect to $\boldsymbol{\theta}$, this definition of $z(f)$ coincides with the definition of the linear least-squares periodogram from (Baluev 2008).

In the majority of practical applications, one of the parameters in $\boldsymbol{\theta}$ is the amplitude K of the periodic variation. This means that

$$\mu(t, \boldsymbol{\theta}, f) = Kh(t, \boldsymbol{\xi}, f), \quad (7)$$

where the vector $\boldsymbol{\xi}$ contains $d - 1$ remaining unknown parameters of the signal. We assume that $\boldsymbol{\xi}$ belongs to some domain Ξ in $d - 1$ dimensions, so that the domain Θ represents the Cartesian product $[0, +\infty) \times \Xi$ or $(-\infty, +\infty) \times \Xi$, and the null domain is the domain of zero amplitude: $\Theta_0 = \{K = 0\} \times \Xi$. For simplicity, let us firstly consider the case when $d_{\mathcal{H}} = 0$ and the hypothesis \mathcal{H} states that the data do not contain anything but the white Gaussian noise. In this case, $\chi_{\mathcal{H}}^2 \equiv \langle x^2 \rangle$ and $\zeta(\boldsymbol{\theta}, f) = \langle x\mu \rangle - \langle \mu^2 \rangle / 2 = \langle xh \rangle K - \langle h^2 \rangle K^2 / 2$ is a quadratic polynomial of K , which can be easily maximized given fixed f and $\boldsymbol{\xi}$. This results in a least-squares estimation $K^* = \langle xh \rangle / \langle h^2 \rangle$, and in the maximum $\max_K \zeta = \eta^2 / 2$, where $\eta(\boldsymbol{\xi}, f)$ represents the new function to be maximized by the remaining parameters. It can be expressed as

$$\eta(\boldsymbol{\xi}, f) = \langle x\psi \rangle, \quad (8)$$

where $\psi(t, \boldsymbol{\xi}, f) = h(t, \boldsymbol{\xi}, f) / \sqrt{\langle h^2 \rangle}$.

A similar result may be obtained for the case when the relation (7) is still valid, but the model $\mu_{\mathcal{H}}$ is no longer empty. It is not hard to check that, if the models μ and $\mu_{\mathcal{H}}$ were orthogonal in the sense that $\langle h\boldsymbol{\varphi}_{\mathcal{H}} \rangle = 0$, the maximum of ζ by K could be expressed exactly in the same way as it was described in the previous paragraph. In the general case the models are not orthogonal, and we introduce the new model function

$$\tilde{h}(t, \boldsymbol{\xi}, f) = h(t, \boldsymbol{\xi}, f) - (\mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}}^{-1} \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, K}) \cdot \boldsymbol{\varphi}_{\mathcal{H}}(t), \quad (9)$$

where $\mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}} = \langle \boldsymbol{\varphi}_{\mathcal{H}} \otimes \boldsymbol{\varphi}_{\mathcal{H}} \rangle$ is the $d_{\mathcal{H}} \times d_{\mathcal{H}}$ Fisher information matrix associated with $\boldsymbol{\theta}_{\mathcal{H}}$, and $\mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, K} = \langle \boldsymbol{\varphi}_{\mathcal{H}} h \rangle$ is the $d_{\mathcal{H}} \times 1$ Fisher information matrix for the parameters $\boldsymbol{\theta}_{\mathcal{H}}$ and K . Since the identity $\langle \tilde{h}\boldsymbol{\varphi}_{\mathcal{H}} \rangle = 0$ holds true, the new model of the signal is orthogonal to the base model. Now \tilde{h} should replace h in the expression for ψ , so that

$$\psi(t, \boldsymbol{\xi}, f) = \frac{h(t, \boldsymbol{\xi}, f) - (\mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}}^{-1} \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, K}) \cdot \boldsymbol{\varphi}_{\mathcal{H}}(t)}{\sqrt{\langle h^2 \rangle - \mathbf{Q}_{K, \boldsymbol{\theta}_{\mathcal{H}}} \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}}^{-1} \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, K}}}. \quad (10)$$

After that, we can directly calculate the quantity η from the equation (8). Finally,

¹ We use here all notation conventions used in (Baluev 2008). For instance, the braces $\{*, *, \dots\}$ denote the association of vectorial or scalar arguments into a single vector, the angular brackets $\langle * \rangle$ denote the summation of the argument over timings t_i with weights $1/\sigma_i^2$, and $\mathbf{x} \otimes \mathbf{y} \equiv \mathbf{x}\mathbf{y}^T$ is the dyadic product of the vectors \mathbf{x} and \mathbf{y}

$$\max_K \zeta = \eta^2/2 \quad (11)$$

with

$$K^* = \eta \sqrt{\langle h^2 \rangle - \mathbf{Q}_{K, \boldsymbol{\theta}_{\mathcal{H}}} \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, K}^{-1} \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, K}}. \quad (12)$$

The best-fitting values of $\boldsymbol{\xi}$ and f correspond to the maximum of η .

Often it might be useful to assume that negative values for K are not allowed. Then we should make a small amendment to the last formulae (11) and (12). Namely, they can be used only for $\eta > 0$, while for $\eta < 0$ we should set $\max_K \zeta = 0$ and $K^* = 0$ with the best-fitting values of $\boldsymbol{\xi}$ and f undefined.

The formulae become more simple for an important practical case when $d_{\mathcal{H}} = 1$ and $\varphi_{\mathcal{H}} \equiv 1$ reflects a free constant offset of the data. In this case, let us first define

$$x_c = \langle x \rangle / \langle 1 \rangle, \quad h_c(\boldsymbol{\xi}, f) = \langle h \rangle / \langle 1 \rangle, \quad D = \langle (h - h_c)^2 \rangle, \quad (13)$$

and then evaluate

$$\eta = \langle (x - x_c)(h - h_c) \rangle / \sqrt{D}, \quad \max_K \zeta = \eta^2/2, \quad K = \eta / \sqrt{D}. \quad (14)$$

Note that the quantity $\langle 1 \rangle$ represents the sum of weights of all observations.

3 APPROXIMATING THE FALSE ALARM PROBABILITY USING THE RICE METHOD

3.1 General introduction to the problem

In this paper we are interested in the false alarm probability (FAP) associated with the observed maximum peak $z_{\max} = \max_{0 \leq f \leq f_{\max}} z(f)$, where f_{\max} is some a priori given maximum frequency. This false alarm probability can be formally defined as follows:

$$\text{FAP}(z_{\max}) = \Pr\{\exists \boldsymbol{\theta} \in \Theta, f \in [0, f_{\max}] : \zeta(\boldsymbol{\theta}, f) > z_{\max}\}, \quad (15)$$

again with the probability operator calculated under the null hypothesis (no actual signal in the data). We can see that to assess the FAP, we should know the distribution function of the maximum values of $\zeta(\boldsymbol{\theta}, f)$. This function represents a real-valued random field defined on a domain of dimension $d + 1$.

One could claim that since the model of the extra signal contains d free parameters and since the quantity $z(f)$ is the logarithm of the likelihood ratio statistic, the distribution of $2z(f)$ (for a fixed f) should tend to the χ^2 distribution with d degrees of freedom, when the number of observations grows. This was assumed, for instance, by Cumming (2004) who considered the case of Keplerian velocity variation with four unknown parameters (plus the period). We must caution the reader that in general this assumption is incorrect because the conditions of the corresponding limiting theorem are not satisfied. The most important reason comes from the fact that the parameters $\boldsymbol{\xi}$ have no physical sense (are undefined) when $K = 0$. Speaking mathematically, they are not identifiable for $K = 0$. The lack of identifiability under the null hypothesis usually destroys the usual asymptotic properties of the likelihood ratio (and χ^2) tests (Dacunha-Castelle & Gassiat 1999). This is because we typically just cannot construct a valid Taylor series of the signal model at $K = 0$, except for rare

special cases. Without that we cannot linearize this model under the null hypothesis, which is critical for the validity of the asymptotic χ^2 distribution. An exception is provided, for instance, by the LS periodogram with the harmonic model of the signal. In this special case, we are able to perform the following re-parametrization: $K \cos(2\pi ft + \lambda) = a \cos(2\pi ft) + b \sin(2\pi ft)$. While that phase λ was a not identifiable at $K = 0$, the new parameters a and b are already identifiable and even linear. In this case, the distribution of each single value of $2z(f)$ is indeed the χ^2 one with two degrees of freedom. Unfortunately any similar trick is not possible for the majority of the other models, even apparently simple ones.

The things get even worse for the more practical case when f is also unknown. In this case the frequency, treated as a new free parameter, is not identifiable (at $K = 0$) even for the LS periodogram. Actually, we may note that non-identifiability of the frequency is the primary obstacle that made the treatment of the significance levels of the LS periodogram so difficult and non-rigorous over decades. The LS periodogram of the noise contains an infinite sequence of similar noisy narrow peaks, but none of them can serve as a reference position for a quadratic Taylor approximation that would be valid in the whole frequency range. If not that obstacle then we could just use the χ^2 distribution with three degrees of freedom (two for a and b plus one for f) to approximate the necessary distribution of the LS periodogram. However, in (Baluev 2008) we managed to deal with this obstacle using the so-called ‘Rice method’, treating the noisy LS periodogram as a random process depending on a real argument f , which was a single non-linear parameter of the model. The case of non-linear periodograms just adds more non-linear arguments of ζ , but the issue of their non-identifiability at $K = 0$ remains qualitatively the same. Therefore, we may try to treat this more general situation using the same or similar method.

Of course, it is hardly possible to derive an exact expression for FAP, but we would be pretty satisfied if we find at least an approximation analogous to what we obtained in our previous works. Namely, we aim to obtain something like

$$\text{FAP}(z_{\max}) \lesssim M(z_{\max}), \quad (16)$$

where M represents simultaneously an upper bound for FAP and its more or less good asymptotic approximation for large z_{\max} (small FAP).

3.2 Basic ideas of the Rice method

The modern comprehensive theory of the Rice method and relevant topics can be found in the reviews (Kratz 2006; Azaïs & Wschebor 2009). Here we present only a very brief extraction of the results that are most useful in our present paper. Suppose we deal with some arbitrary random process or field $Z(\mathbf{x})$ and we need to find the probability that its maximum (within some domain $\mathbf{x} \in \mathbb{X}$) will lie beyond a specified level $Z(\mathbf{x}) = z$. In our signal detection task this probability is equal to $\text{FAP}(z)$, and this is obviously a complementary probability to the distribution function of the maximum of Z . The general Rice method to estimate these thing is based on two main points. First, we should construct some derived integer random variable $\mathcal{N}(z)$, such that the

event $\mathcal{N}(z) = 0$ is equivalent (or almost equivalent) to the event $\{Z(\mathbf{x}) < z \forall \mathbf{x} \in \mathbb{X}\}$, and the event $\mathcal{N}(z) \geq 1$ is (almost) equivalent to $\{\exists \mathbf{x} \in \mathbb{X} : Z(\mathbf{x}) > z\}$. The boundary event when $Z(\mathbf{x}) \leq z$ everywhere in \mathbb{X} and there is only one or a few \mathbf{x} such that $Z(\mathbf{x}) = z$ should usually correspond to $\mathcal{N}(z) = 1$. The word “almost” refers here only to the effects at the boundary of \mathbb{X} (the boundary maxima); if we somehow knew for sure that boundary maxima are impossible than this word can be just omitted.

For random processes a good choice for \mathcal{N} is the number of up-crossings of the specified level; i.e. the number of points x such that $Z(x) = z$ and $Z'(x) > 0$. For random fields the term of up-crossing is meaningless, and in this case we choose $\mathcal{N}(z)$ to be the number of local maxima beyond z (and inside \mathbb{X}), that is the number of points \mathbf{x} where $Z > z$, $Z' = 0$, and Z'' is negative-definite.

Given such a counter variable \mathcal{N} , we can estimate the required false alarm probability, i.e. the probability for $Z(\mathbf{x})$ to exceed a given level z somewhere in \mathbb{X} , as

$$\text{FAP}(z) \lesssim M(z) = M_{\text{boundary}}(z) + \tau(z), \quad \tau(z) = \mathbb{E}\mathcal{N}(z). \quad (17)$$

Here the term M_{boundary} refers to the maxima attained at the boundary of \mathbb{X} ; it may or may not be neglected, depending on other conditions of the task. We will discuss it in detail later. The primary term is $\tau(z)$, which is equal to the mathematical expectation of the selected counter. This formula is basically the same as (16) with concretized function $M(z)$.

The second point of the Rice method is the generalized Rice formula for τ . For the random processes we have

$$\tau(z) = \int_{\mathbb{X}} \mathbb{E}([Z'(x)]_+ | Z(x) = z) p_Z(z) dx, \quad (18)$$

where the functions p stand for the probability density functions of the quantity $Z(x)$ shown as indices (this p also depends on x), and $[a]_+ = \max(0, a)$. If necessary, the expression (18) can be obviously rewritten in terms of the joint distribution of $Z(x)$ and $Z'(x)$, see (Baluev 2008). The name “Rice method” and “Rice formula” are after Rice (1944), who constructed his original Rice formula for the stationary Gaussian random process.

In this paper we will use the generalized Rice formula for random fields. Actually, now we have even three formulae of that type. The first one is introduced in (Azaïs & Delmas 2002); it can be written down as

$$\tau(z) = \int_z^\infty dZ \int_{\mathbb{X}} \mathbb{E}(\det[Z'']_- | Z' = 0; Z) p_{Z, Z'}(Z, 0) d\mathbf{x}, \quad (19)$$

where $\det[Z'']_-$ is equal to $|\det Z''|$ when Z'' is negative-definite, and zero otherwise. This formula is generally similar to (18), although considerably more complicated. Azaïs & Wschebor (2009) introduced in their Chapter 8 a variation of (19) with $\det[Z'']_-$ replaced by $|\det Z''|$. Such replacement obviously somewhat increases the right-hand side of (19), keeping its upper-limit property, but making the computations a bit more easy. It counts *all* the critical points of $Z(\mathbf{x})$ above z , not just the local maxima. However, both these formulae are usually too difficult for computations, and the formula that is typically used in practice contains

just the “naked” $\det Z''$ instead of $|\det Z''|$ or $\det[Z'']_-$. Such a formula gives the mathematical expectation of the Euler-Poincaré characteristic (EPC) of the level-section set $\{\mathbf{x} \in \mathbb{X} : Z(\mathbf{x}) > z\}$ (which is also called as the “excursion set”).

Unfortunately, the quantity $\mathbb{E}(\text{EPC})$ does not strictly retain the upper-limit property of $\mathbb{E}\mathcal{N}$. However, it is known (at least for Gaussian fields, see e.g. Chapter 8 by Azaïs & Wschebor (2009)) that for large levels z the quantities $\mathbb{E}\mathcal{N}$ and $\mathbb{E}(\text{EPC})$ are asymptotically equivalent, and their difference decreases rather quickly (we will detail this below). This is because beyond a large z all critical points of Z are local maxima with almost unit probability, so the relevant excursion set represents a number of (filled) ellipsoids encompassing the positions of these maxima. Each such a filled ellipsoid has $\text{EPC} = 1$, and thus $\text{EPC} \simeq \mathcal{N}$ for large z . All this means that we can typically use $\mathbb{E}(\text{EPC})$ as a good practical approximation for τ . Even if $\mathbb{E}(\text{EPC})$ does not provide an entirely strict upper bound, the relevant errors usually appear negligible for practical levels of z . Of course we must admit that “usually” or “typically” is not the same as “always”, but nonetheless this approximation appears quite satisfactory in the examples considered below in the paper, as well as in a few other cases that we prepare for a future publication.

The Rice method usually provides good practical accuracy, so that the mentioned upper bound (16) appears close to the actual value of FAP, at least for practically important case of small FAP levels. The Rice method does not belong to widely-known methods, because it is not mentioned in a typical handbook on mathematical statistics. Therefore, its usage in applications (e.g. in astronomy) is rare. However, rare does not mean absent: we found that some variant of this method was applied by Bardeen et al. (1986) to study cosmological density fluctuations, which were modelled by a Gaussian random field.

3.3 Applying the Rice method to non-linear periodograms

Mathematically, the condition of $\max \zeta \leq z$ is equivalent to that of $\max |\eta| \leq \sqrt{2z}$ (case of arbitrary K) or $\max \eta \leq \sqrt{2z}$ (case of $K \geq 0$). Therefore, we need to calculate the distribution of the maximum values of the random function $\eta(\boldsymbol{\xi}, f)$ to estimate the FAP. In this subsection we limit ourselves by the single-sided case $K \geq 0$, bearing in mind that to obtain the formulae for the case of arbitrary K we need to double the right-hand side of (17), because then we need to honour the maxima of η above $\sqrt{2z}$ as well as its minima below $-\sqrt{2z}$, which are entirely analogous.

The random field $\eta(\boldsymbol{\xi}, f)$ possesses quite simple statistical properties. From the definition (8) it clearly follows that if the noise in our observations is Gaussian, η represents a Gaussian random field. It is easy to check from (8) and (10) that $\mathbb{E}\eta \equiv 0$ and the variance $\mathbb{D}\eta \equiv 1$. This places us in the framework of Theorem 1 by Azaïs & Delmas (2002). In this case we can use (17) with

$$\tau(z) \approx \mathbb{E}(\text{EPC}) = \sum_{j=0}^{[n/2]} a_j P_{n+1-2j}(z). \quad (20)$$

In this relation, the integer n represents the number of free

arguments of η . It is equal to $\dim \xi = d - 1$ or $\dim \xi + 1 = d$, depending on whether we consider the case of fixed or free frequency f . Below we will only consider the more practical second case with $n = d$, but in order to avoid misunderstandings we prefer to keep different notations for n and d . The notation $[*]$ stands for the integer part of the argument. The functions $P_k(z)$ represent the tail probability associated with the χ^2 distribution with k degrees of freedom:

$$P_k(z) = \frac{1}{\Gamma(k/2)} \int_z^\infty x^{k/2-1} e^{-x} dx. \quad (21)$$

When $z \rightarrow \infty$, the error of the approximation in (20) has the decrease rate quicker than $e^{-(1+\delta)z}$ with some positive δ , while τ itself typically decreases as $e^{-z} z^{(n-1)/2}$. This means that the relevant *relative* error decreases quicker than $e^{-z\delta}/z^{(n-1)/2}$.

We can see that the expression (17) involves a linear combination of χ^2 tails with different numbers of degrees of freedom. However, the sum of the coefficients a_j is not necessary unit, so that the expansion in (17) is not a mixture of distributions in the rigorous meaning of this notion. For large z , the first term with $P_{n+1} \sim z^{(n-1)/2} e^{-z}$ dominates, whereas the relative magnitudes of the remaining terms decrease as $\sim 1/z^j$.

The calculation of the coefficients a_j represents a large technical difficulty. These coefficients are proportional to the quantities k_{2j} introduced in Theorem 1 by Azaïs & Delmas (2002). We altered the original coefficients k_{2j} in order to have the functions P_k explicitly in the sum (20).

Let us denote the variance-covariance matrix of the gradient of η as \mathbf{G} (it is denoted as Λ by Azaïs & Delmas (2002)). It can be easily calculated from the eq. (8). The part of the matrix \mathbf{G} corresponding to only the parameters ξ is equal to

$$\left\langle \frac{\partial \psi}{\partial \xi} \otimes \frac{\partial \psi}{\partial \xi} \right\rangle, \quad (22)$$

and the remaining elements due to the frequency parameter can be expressed in an entirely analogous manner.

At first, let us consider a simplified situation when \mathbf{G} does not depend on ξ and f . Then we can use the proposition “a” of Theorem 1 in (Azaïs & Delmas 2002). We find that in this case the coefficients a_j (or k_{2j}) are proportional to the coefficients of the Hermite polynomials H_n (Korn & Korn 1968, §21.7), so that after some elementary transformations we have

$$\tau \approx A F_n(z) \quad (23)$$

with

$$A = \frac{\sqrt{\det \mathbf{G}}}{2\pi^{(n+1)/2}} \text{Vol}(\Xi) f_{\max},$$

$$F_n(z) = \int_z^\infty H_n(\sqrt{x}) e^{-x} \frac{dx}{\sqrt{x}} = e^{-z} H_{n-1}(\sqrt{z}). \quad (24)$$

The last equality in (24) can be just checked by direct differentiation or it can be derived “honestly” using the Rodrigues representation for H_n . Notice that H_n are normalised here so that their highest coefficients are equal to unit and the weighting function is e^{-x^2} .

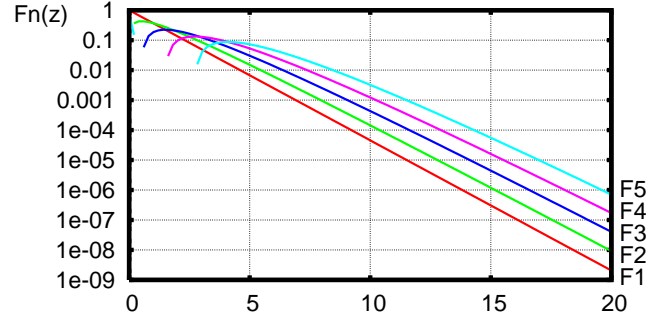


Figure 1. The graphs of several functions $F_n(z)$.

Let us write down a few of the functions F_n :

$$F_1(z) = e^{-z}, \quad F_2(z) = e^{-z} \sqrt{z}, \quad F_3(z) = e^{-z} \left(z - \frac{1}{2} \right),$$

$$F_4(z) = e^{-z} \left(z - \frac{3}{2} \right) \sqrt{z}, \quad F_5(z) = e^{-z} \left(z^2 - 3z + \frac{3}{4} \right). \quad (25)$$

We plot the graphs of these functions in Fig. 1. Also, a notable recursive relation $F_{n+1}(z) = -F'_n(z) \sqrt{z}$ can be used.

When \mathbf{G} is non-constant, which is a more practical case, we can easily calculate only the primary term in (20). We have in this case

$$\tau \approx A e^{-z} z^{(n-1)/2} \quad (26)$$

with the relative error of $\sim 1/z$, which is worse than the error of (20). Here

$$A = \frac{1}{2\pi^{(n+1)/2}} \int_0^{f_{\max}} df \int_{\Xi} \sqrt{\det \mathbf{G}} d\xi, \quad (27)$$

which is an evident generalization of A from (24).

The remaining z -power terms are different from the case of constant \mathbf{G} , and usually they are very hard to evaluate, because they involve conditional covariances of the *second-order* derivatives of η in quite unpleasant combinations (see proposition “b” of Theorem 1 by Azaïs & Delmas 2002). It might be noticed that for a simple case $n = 1$ we have only one term in (20).

3.4 The role of the boundary effects

The quantity τ on itself does not yet provide a closed solution of the problem. According to (17), we need to assess a similar term, which is related to the number of the so-called “boundary maxima” (which are the local maxima of the random field restricted to the domain \mathbb{X} boundary).

This term takes into account the situation when all local maxima in the domain interior appear smaller than some boundary maximum. It can be calculated using very similarly to the term τ , we just need to consider the restriction of our task to this boundary and apply the Rice method in a recursive manner. We should approximate the quantity M_{boundary} by a formula similar to (17) and (20), but with n replaced by $n - 1$. Hence, its relative contribution decreases for large z , but at a rather slow rate of $\sim 1/\sqrt{z}$.

We however must take care of one small but rather important thing. When we restrict our field η to the boundary of our domain, a maximum at the boundary does not

necessarily represent a good candidate for the global maximum. Whether a particular boundary maximum is “good” or “bad”, depends on the sign of the derivative of η in the direction of the outward normal to the boundary (i.e., the projection of the gradient η' to this normal). If it is negative, we can definitely find larger values of η when stepping from the boundary inwards. Thus, such a boundary maximum can never provide the global maximum of η , so we call it “bad”. When counting the boundary maxima, we must filter out all “bad” ones, keeping only the “good”, which offer positive outward derivative. Mathematically, all this requests from us to replace in (19) the gradient Z' and the Hessian Z'' by their projections to the tangent plane to the boundary, (denote them, say, Z'_{\parallel} and Z''_{\parallel}), and consider the operator \mathbb{E} and the probability density $p_{Z,Z'_{\parallel}}$ conditionally to an additional constraint $Z'_{\perp} > 0$ (where index \perp means the projection on the outward normal to the boundary surface).

In practice this usually just means that we need to halve (precisely or approximately) the estimated number of the boundary maxima, τ_{boundary} . We must admit that this issue has not been investigated in the literature with enough details. Azais & Delmas (2002) prove this “1/2-rule” under certain restrictive assumptions, among which the most important is the requirement of constant \mathbf{G} . From their Theorem 3 it follows, basically, that

$$M_{\text{boundary}} = \frac{1}{2} \tau_{\text{boundary}} + \dots, \quad (28)$$

where τ_{boundary} can be evaluated in essentially the same manner as τ , considering the restriction of the task to the boundary surface (which implies, in particular, a decrease in n), and under “...” we mean here some terms having faster decrease rate than τ_{boundary} . Although Azais & Delmas (2002) leave the case when $\mathbf{G} \neq \text{const}$ aside, after investigation of their detailed proofs, we do not find an obstacle in generalizing their single-term asymptotic formula for M_{boundary} to a more general case with $\mathbf{G} \neq \text{const}$. Moreover, we find that the main neglected term contained in “...” of (28) has the relative magnitude of $\sim 1/z$.

Since in the general case we anyway keep only the greatest terms in τ and similar quantities, we can use a two-term approximation like

$$\text{FAP}(z) \lesssim M(z) \simeq \left(A z^{(n-1)/2} + \frac{1}{2} A_{\text{boundary}} z^{n/2-1} \right) e^{-z}. \quad (29)$$

The principal error of the right-most expression has the relative magnitude of $\sim 1/z$ and is due to the omitted terms in τ , while the omitted terms in τ_{boundary} are of even a smaller order $\sim 1/z^{3/2}$.

In more complicated cases, the boundary itself may be non-smooth due to “sub-boundaries” of smaller dimension (edges, vertexes), which will generate extra terms in M_{boundary} . It is rather difficult to formulate a simple and general receipt of how to deal them with, because the geometry of the boundary might be quite complicated in general. However, later we explain this procedure on a concrete example of the von Mises periodogram, when the parametric domain is a rectangle.

4 EVALUATING THE COEFFICIENT A

4.1 Assuming the uniform phase coverage

It would be useful to construct the analytic expressions of the FAP for the case when our observations are distributed approximately uniformly in time. We assume that the timings t_i , when they are phased to some frequency f , cover the relevant phase more or less uniformly. In this case, we can approximate the summation $\langle * \rangle$ over the time series by means of integration over its time span $[t_1, t_2]$. Saying it more accurately, we approximate the time-series average $\langle * \rangle / \langle 1 \rangle$ by the integral average over the time span. Moreover, the periodic character of the model μ usually allows for this integration to be performed over a single period only. The periodicity of h implies that $h(t, \xi, f) = g(2\pi ft + \lambda, \nu)$, where the function $g(x, \nu)$ is 2π -periodic in x and λ is the phase parameter. The remaining $d - 2$ parameters form the vector ν . This means that $\Xi = [0, 2\pi] \times \Upsilon$ and $\Theta = [0, +\infty) \times [0, 2\pi] \times \Upsilon$, where Υ is some $(d - 2)$ -dimensional domain of parameters ν . Now we can approximate, for instance, the mean value of h as

$$\langle h \rangle \approx \frac{w}{P} \int_0^P h(t, \xi, f) dt = \frac{w}{2\pi} \int_0^{2\pi} g(x, \nu) dx = w\bar{g}, \quad (30)$$

where $P = 1/f$ is the period of the signal, $w = \langle 1 \rangle$ is the sum of the weights and the overline denotes the continuous averaging over the periodic variable x . Note that usually we deal with the case when the base model $\mu_{\mathcal{H}}$ incorporates a constant term. According to (10), this means that we should first subtract the constant $\langle g \rangle / w \approx \bar{g}$ from the model of the signal. We will assume that g was already centered.

The same arguments lead to the equality $\langle h^2 \rangle \approx w\bar{g}^2$. The latter expression does not depend on f and λ , so we can write down the derivatives of ψ as

$$\frac{\partial \psi}{\partial f} \approx \frac{2\pi t g'_x}{\sqrt{wg^2}}, \quad \frac{\partial \psi}{\partial \lambda} \approx \frac{g'_x}{\sqrt{wg^2}}, \quad \frac{\partial \psi}{\partial \nu_i} \approx \frac{g'_{\nu_i}}{\sqrt{wg^2}} - \frac{g \overline{g g'_{\nu_i}}}{\sqrt{wg^2}^3}, \quad (31)$$

where the function g and its derivatives $g'_x = \partial g / \partial x$ and $g'_{\nu_i} = \partial g / \partial \nu_i$ outside the averaging are calculated for $x = 2\pi ft + \lambda$. This allows us to calculate the elements of the matrix \mathbf{G} :

$$\begin{aligned} \left\langle \left(\frac{\partial \psi}{\partial \lambda} \right)^2 \right\rangle &\approx q \equiv \frac{\overline{g'^2_x}}{g^2}, \\ \left\langle \frac{\partial \psi}{\partial \lambda} \frac{\partial \psi}{\partial \nu_i} \right\rangle &\approx v_i \equiv \frac{\overline{g'_x g'_{\nu_i}}}{g^2} - \frac{\overline{g g'_x g'_{\nu_i}}}{g^2}, \\ \left\langle \frac{\partial \psi}{\partial \nu_i} \frac{\partial \psi}{\partial \nu_j} \right\rangle &\approx V_{ij} \equiv \frac{\overline{g'_{\nu_i} g'_{\nu_j}}}{g^2} - \frac{\overline{g g'_{\nu_i} g'_{\nu_j}}}{g^2}, \end{aligned} \quad (32)$$

which do not depend on the frequency and phase. When calculating the elements of the matrix \mathbf{G} , we also deal with summations like, for instance,

$$\left\langle \left(\frac{\partial \psi}{\partial f} \right)^2 \right\rangle \approx \frac{4\pi^2}{wg^2} \left\langle t^2 g'^2_x (2\pi ft + \lambda) \right\rangle. \quad (33)$$

The P -periodic function g'^2_x can be expanded in the Fourier series with the constant term being equal to $\overline{g'^2_x}$. If t_i span a large enough number of the periods approximately uni-

formly, the summation in (33) averages out all Fourier harmonics, except for the constant term, which results in

$$\left\langle \left(\frac{\partial \psi}{\partial f} \right)^2 \right\rangle \approx 4\pi^2 \bar{t}^2 q. \quad (34)$$

Here, the line over t^2 denotes the weighted averaging of the squared timings: $\bar{t}^2 = \langle t^2 \rangle / \langle 1 \rangle$, which can be easily evaluated directly (without approximating it by a continuous integral). Similar arguments lead to

$$\left\langle \frac{\partial \psi}{\partial f} \frac{\partial \psi}{\partial \lambda} \right\rangle \approx 2\pi \bar{t} q, \quad \left\langle \frac{\partial \psi}{\partial f} \frac{\partial \psi}{\partial \nu_i} \right\rangle \approx 2\pi \bar{t} v_i, \quad (35)$$

where \bar{t} is the weighted average of t_i . Thus, the full matrix \mathbf{G} can be written in the following block form:

$$\mathbf{G} \approx \begin{pmatrix} 4\pi^2 \bar{t}^2 q & 2\pi \bar{t} q & 2\pi \bar{t} \mathbf{v}^T \\ 2\pi \bar{t} q & q & \mathbf{v}^T \\ 2\pi \bar{t} \mathbf{v} & \mathbf{v} & \mathbf{V} \end{pmatrix}, \quad (36)$$

where the elements of the vector \mathbf{v} and those of the matrix \mathbf{V} are defined in (32). In this approximation, the matrix \mathbf{G} only depends on the parameters $\boldsymbol{\nu}$. Using simple elementary transformations of \mathbf{G} we can finally obtain that

$$\det \mathbf{G} \approx \pi q^2 T_{\text{eff}}^2 \det \mathbf{R}, \quad (37)$$

where $\mathbf{R} = \mathbf{V} - \mathbf{v} \otimes \mathbf{v} / q$ (that is, $R_{ij} = V_{ij} - v_i v_j / q$) and $T_{\text{eff}} = \sqrt{4\pi (\bar{t}^2 - \bar{t}^2)}$ is the effective length of the time series, as it was defined in (Baluev 2008). Integrating (27) over λ and substituting (37), we can write down:

$$A \approx \frac{W}{\pi^{n/2-1}} \int_{\Upsilon} q \sqrt{\det \mathbf{R}} d\boldsymbol{\nu}, \quad W = f_{\text{max}} T_{\text{eff}}. \quad (38)$$

The factor A can now be substituted in (23) for the use in (17) and (16). In the degenerate case $n = 2$, we have $\dim \boldsymbol{\nu} = 0$ and put $\det \mathbf{R} = 1$ by definition.

The main advantage of this method of calculation of A is that the result depends on a particular time series only via the single quantity T_{eff} . Given the model g and the domain Υ , we can evaluate an approximation for A once and then use it for all time series, substituting the proper values of f_{max} and T_{eff} . For the LS periodogram, for instance, we have $A \approx W$, which is in full agreement with (Baluev 2008). The main disadvantage is that this approximation may have insufficient accuracy in practice.

When deriving this approximation for A , we assumed the uniform phase coverage for all frequencies f in the scan range. When the original time series is not uniform, this assumption may become invalid at some frequencies f , corresponding to periodic leakage patterns of t_i (and including also the zero frequency). However, these perturbations can usually appear only inside very short frequency segments ($\Delta f \sim 1/T_{\text{eff}}$) around the leakage frequencies; for the most values of f in the range $[0, f_{\text{max}}]$ the phase coverage is still uniform. But our formula for A in (27) involves an integration over the wide frequency band $\Delta f \sim W/T_{\text{eff}}$ with $W \gg 1$, and hence the perturbations of its integrand have almost no effect on the result, because they are limited by so short frequency segments. This means that the accuracy of our approximation for A does not significantly degrades even for non-uniform time series. For sinusoidal signals this was already demonstrated in (Baluev 2008), where we have

shown that even ultimately strong aliasing has only a negligible effect on the resulting A , when $W \gtrsim 10$.

In the case of non-sinusoidal signals, an important source of the errors of the approximation (38) comes from another side. If the signal model contains some quickly varying structures, e.g. narrow peaks, the observations may cover these structures with insufficient sampling, so the resulting approximation for A may appear poor even when t_i are perfectly uniform. This effect is important for the von Mises periodogram below, for example.

4.2 Evaluating A directly

The direct evaluation of the factor A by means of substitution of (22) to (27) involves rather unpleasant manipulations with huge formulae, especially when we work in general terms of Section 2. To simplify them, let us write down the gradient of the model h over the compound vector of all non-linear parameters $\boldsymbol{\omega} = \{f, \lambda, \boldsymbol{\nu}\}$:

$$\boldsymbol{\gamma} = \frac{\partial h}{\partial \boldsymbol{\omega}} = \{2\pi t g'_x, g'_x, g'_\nu\}. \quad (39)$$

Then define

$$\begin{aligned} \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}} &= \langle \boldsymbol{\varphi}_{\mathcal{H}} \otimes \boldsymbol{\varphi}_{\mathcal{H}} \rangle, \\ \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, K} &= \langle \boldsymbol{\varphi}_{\mathcal{H}} g \rangle, \\ \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\omega}} &= \langle \boldsymbol{\varphi}_{\mathcal{H}} \otimes \boldsymbol{\gamma} \rangle, \\ \mathbf{T} &= \langle \boldsymbol{\gamma} \otimes \boldsymbol{\gamma} \rangle - \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\omega}}^T \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}}^{-1} \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\omega}}, \\ \mathbf{y} &= \langle g \boldsymbol{\gamma} \rangle - \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\omega}}^T \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}}^{-1} \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, K}, \\ D &= \langle g^2 \rangle - \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, K}^T \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}}^{-1} \mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, K}, \end{aligned} \quad (40)$$

where $\boldsymbol{\varphi}_{\mathcal{H}}$ is the functional base of the linear null model (3). Finally,

$$\mathbf{G} = \frac{\mathbf{T}}{D} - \frac{\mathbf{y} \otimes \mathbf{y}}{D^2}. \quad (41)$$

We may also offer another evaluation sequence. Let us construct the full Fisher information matrix of all the parameters involved $(\boldsymbol{\theta}_{\mathcal{H}}, K, \boldsymbol{\omega})$:

$$\mathbf{Q} = \begin{pmatrix} \boldsymbol{\varphi}_{\mathcal{H}} \otimes \boldsymbol{\varphi}_{\mathcal{H}} & \boldsymbol{\varphi}_{\mathcal{H}} g & \boldsymbol{\varphi}_{\mathcal{H}} \otimes \boldsymbol{\gamma} \\ g \boldsymbol{\varphi}_{\mathcal{H}}^T & g^2 & g \boldsymbol{\gamma}^T \\ \boldsymbol{\gamma} \otimes \boldsymbol{\varphi}_{\mathcal{H}} & \boldsymbol{\gamma} g & \boldsymbol{\gamma} \otimes \boldsymbol{\gamma} \end{pmatrix}. \quad (42)$$

Please notice that the triangle braces, standing for the weighted summation over the time series, are still here. Now let us apply the Cholesky decomposition: $\mathbf{Q} = \mathbf{L} \mathbf{L}^T$ with \mathbf{L} being a lower-triangular matrix. Then write down \mathbf{L} in the same block form as \mathbf{Q} in (42):

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}} & 0 & 0 \\ \mathbf{l}_{K, \boldsymbol{\theta}_{\mathcal{H}}}^T & l_{K, K} & 0 \\ \mathbf{L}_{\boldsymbol{\omega}, \boldsymbol{\theta}_{\mathcal{H}}} & \mathbf{l}_{\boldsymbol{\omega}, K} & \mathbf{L}_{\boldsymbol{\omega}, \boldsymbol{\omega}} \end{pmatrix}. \quad (43)$$

Notice that $\mathbf{l}_{K, \boldsymbol{\theta}_{\mathcal{H}}}$ and $\mathbf{l}_{\boldsymbol{\omega}, K}$ are vectors, and $l_{K, K}$ is a scalar. Obviously, $\mathbf{L}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}}$ is a lower-triangular matrix of the Cholesky decomposition for $\mathbf{Q}_{\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\theta}_{\mathcal{H}}}$, and $\mathbf{L}_{\boldsymbol{\omega}, \boldsymbol{\omega}}$ is another lower-triangular matrix. From the matrix \mathbf{L} definition and from (41) we can easily derive two remarkable relations:

$$\mathbf{G} = \frac{\mathbf{L}_{\boldsymbol{\omega}, \boldsymbol{\omega}} \mathbf{L}_{\boldsymbol{\omega}, \boldsymbol{\omega}}^T}{l_{K, K}^2}, \quad D = l_{K, K}^2. \quad (44)$$

Moreover, it is clear that $\sqrt{\det \mathbf{G}} = \det \mathbf{L}_{\boldsymbol{\omega}, \boldsymbol{\omega}} / l_{K, K}^d$. Therefore, to find the integrand in (27) we just need to calculate

$\det \mathbf{L}_{\omega, \omega}$, which is simply equal to the product of its diagonal elements, and then divide the result by $l_{K,K}^d$.

Therefore, the final procedure to evaluate $\sqrt{\det \mathbf{G}}$ is as follows.

(i) Evaluate \mathbf{Q} using its definition (42). Its size should be $d_{\mathcal{H}} + d + 1$. Note that it is important to preserve the ordering of the parameters as $\theta_{\mathcal{H}}, K, \omega$, while the ordering inside $\theta_{\mathcal{H}}$ and inside ω is not important.

(ii) Perform the Cholesky decomposition of \mathbf{Q} . It is very quick and numerically stable procedure.

(iii) On the basis of only diagonal elements of the resulting Cholesky matrix \mathbf{L} , construct the combination $(\prod_{i=d_{\mathcal{H}}+2}^{d_{\mathcal{H}}+d+1} l_{ii}) / (l_{d_{\mathcal{H}}+1, d_{\mathcal{H}}+1})^d$. It is equal to what we seek.

The quantity $\sqrt{\det \mathbf{G}}$ must be further numerically integrated over the parameters ω , according to (27). Note that this includes the integration over the frequency f and over the phase λ (which is now non-trivial).

Now, let us limit ourself to the most important practical case when the null model involves only a free constant term: $d_{\mathcal{H}} = 1$, $\varphi_{\mathcal{H}} \equiv 1$. In this case the matrix \mathbf{Q} is considerably simplified:

$$\mathbf{Q} = \begin{pmatrix} 1 & g & \gamma^T \\ g & g^2 & g\gamma^T \\ \gamma & g\gamma & \gamma \otimes \gamma \end{pmatrix}. \quad (45)$$

The round-off errors may destroy the positive-definiteness of \mathbf{Q} , which is critical for the Cholesky decomposition. To reduce this effect, the computational sequence can be transformed to something similar to the formulae (13) and (14). This will be some hybrid approach to evaluate $\sqrt{\det \mathbf{G}}$ between the two ones that we have already discussed. Namely, first we should center the functions involved:

$$g_c = \langle g \rangle / \langle 1 \rangle, \quad \gamma_c = \langle \gamma \rangle / \langle 1 \rangle, \\ \tilde{g} = g - g_c, \quad \tilde{\gamma} = \gamma - \gamma_c. \quad (46)$$

After that, we need to evaluate

$$D = \langle \tilde{g}^2 \rangle, \\ \mathbf{y} = \langle \tilde{g} \tilde{\gamma} \rangle, \\ \beta = \tilde{\gamma} - \tilde{g} \mathbf{y} / D, \quad (47)$$

which imply that

$$\mathbf{G} = \langle \beta \otimes \beta \rangle / D. \quad (48)$$

We have no need to evaluate \mathbf{G} itself. Instead, we may perform the Cholesky decomposition of $\langle \beta \otimes \beta \rangle$. Dividing the product of the diagonal elements of the resulting Cholesky matrix by $D^{d/2}$, we obtain $\sqrt{\det \mathbf{G}}$.

The obvious advantage of the direct method to evaluate the factor A is that it allows for high accuracy limited by only round-off and numerical integration errors, not relying on any approximating assumptions. The disadvantage is that it is considerably more slow than in Section 4.1, although in practice the relevant computation time should be comparable to the time of a single evaluation of the periodogram itself.² Therefore, this is still much faster than e.g. Monte Carlo simulation, where this periodogram has to be

re-evaluated thousands of times before we reach a reliable FAP estimation.

5 UNKNOWN NOISE LEVEL

In Section 2, we have assumed that the standard errors σ_i of observations are known a priori. In practice we often do not know them with enough precision. A commonly used model is given by $\sigma_i^2 = \kappa / w_i$, where the quantities w_i determine the weighting pattern of the time series and the factor κ remains unconstrained a priori. Similar problem was considered in the paper (Baluev 2009a). In this work, the model $\sigma_i^2 = \sigma_{\text{meas},i}^2 + \sigma_{\star}^2$ was considered with $\sigma_{\text{meas},i}^2$ being the ‘internal’ measurements variances, known a priori, and parameter σ_{\star}^2 being the unconstrained variance of the extra ‘jitter’. In these cases, we cannot calculate the least-squares periodogram $z(f)$, since we cannot calculate the values of the χ^2 functions themselves.

The general approach for solving such problems is based on the likelihood ratio test. The logarithm of the likelihood function for our observations, which are contaminated by random mutually independent Gaussian errors, may be written down (specifically for the hypothesis \mathcal{H} and \mathcal{K}) as

$$\ln \mathcal{L}_{\mathcal{H}, \mathcal{K}} = -\frac{1}{2} \sum_{i=1}^N \left[\frac{(x_i - \mu_{\mathcal{H}, \mathcal{K}}(t_i))^2}{\sigma_i^2(\mathbf{p})} + \ln \sigma_i^2(\mathbf{p}) \right] + \text{const.} \quad (49)$$

Here the full variances σ_i^2 depend on the extra parameters \mathbf{p} , which should be estimated from the data together with the usual parameters $(\theta_{\mathcal{H}}, \theta, f)$ of the model curve. These estimations are obtained in result of maximizing the corresponding likelihood function over all of the parameters to be estimated. After that, we could construct the logarithm of the likelihood ratio statistic as the maximum (over the frequency f) of the likelihood ratio periodogram

$$Z(f) = \max_{\mathbf{p}, \theta_{\mathcal{K}}} \ln \mathcal{L}_{\mathcal{K}}(\mathbf{p}, \theta_{\mathcal{K}}, f) - \max_{\mathbf{p}, \theta_{\mathcal{H}}} \ln \mathcal{L}_{\mathcal{H}}(\mathbf{p}, \theta_{\mathcal{H}}). \quad (50)$$

This function may give the basis for signal detection in the general framework, when \mathbf{p} is not known a priori. However, for the aims of reduction of the statistical bias in \mathbf{p} , it is better to use the following modifications of the likelihood functions and of the likelihood ratio periodogram:

$$\ln \tilde{\mathcal{L}}_{\mathcal{H}, \mathcal{K}} = -\frac{1}{2} \sum_{i=1}^N \left[\frac{(x_i - \mu_{\mathcal{H}, \mathcal{K}}(t_i))^2}{\gamma_{\mathcal{H}, \mathcal{K}} \sigma_i^2(\mathbf{p})} + \ln \sigma_i^2(\mathbf{p}) \right], \\ \tilde{Z}(f) = \gamma_{\mathcal{K}} \left[\max_{\mathbf{p}, \theta_{\mathcal{K}}} \ln \tilde{\mathcal{L}}_{\mathcal{K}} - \max_{\mathbf{p}, \theta_{\mathcal{H}}} \ln \tilde{\mathcal{L}}_{\mathcal{H}} \right] + \frac{N_{\mathcal{K}}}{2} \ln \frac{N_{\mathcal{H}}}{N_{\mathcal{K}}}, \quad (51)$$

where $N_{\mathcal{H}} = N - d_{\mathcal{H}}$, $N_{\mathcal{K}} = N - d_{\mathcal{K}}$, and $\gamma_{\mathcal{H}, \mathcal{K}} = N_{\mathcal{H}, \mathcal{K}} / N$. This modification was discussed in details in the paper (Baluev 2009a).

For the popular practical case $\sigma_i^2 = \kappa / w_i$ with w_i known a priori, the periodogram $\tilde{Z}(f)$ represents a direct extension of the periodogram $z_3(f)$ from (Baluev 2008). It can be constructed now as the maximum (over θ) of the non-linear

² We will typically evaluate this periodogram on some multidimensional grid in the space of all the parameters ω , not just on a

frequency grid like in the LS case. In terms of the computational demands, this procedure is roughly equivalent to the numerical integration over the same space.

function

$$\zeta_3(\boldsymbol{\theta}, f) = -\frac{N_K}{2} \ln \left[1 - \frac{2\zeta(\boldsymbol{\theta}, f)}{\min_{\boldsymbol{\theta}_H} \chi_{\mathcal{H}}^2(\boldsymbol{\theta}_H)} \right]. \quad (52)$$

Note that the ratio $\zeta/\chi_{\mathcal{H}}^2$ is independent of κ , so that ζ_3 can be calculated regardless the factor κ is unknown.

To assess the statistical significance of the peaks on the likelihood ratio periodogram, we need to know the distributions of the maxima of $\tilde{Z}(f)$, as previousl. Now we cannot just apply the results from the previous sections directly. However, we may use the asymptotic large-sample properties of the likelihood function.

To do this we have to assume that the new parameters \mathbf{p} do not introduce any extra peculiarities like e.g. extra non-identifiability under \mathcal{H} . This is normally true. The calculations involving quadratic Taylor expansion of the likelihood function near the point $K = 0$, show that in the asymptotic $N \rightarrow \infty$ approximation for the quantity η looks exactly the same as its linear-case definition in (8) and (10). This means that the joint distribution of η and of its gradient remains asymptotically the same as in the genuine linear case. Consequently, all the theory of Section 3 remains valid for the likelihood-ratio periodograms as an asymptotic approximation for $N \rightarrow \infty$. The same holds for the modified likelihood functions and for the modified likelihood ratio periodogram (51), since this modification does not introduce any change in the asymptotic behaviour.

For the periodogram $z_3(f)$ and linear models of the signal, the approximations to the periodograms distributions for arbitrary N are given in (Baluev 2008). When $N \rightarrow \infty$, these approximations indeed rapidly converge to those obtained for the original least-squares periodogram. This provides an independent confirmation of the arguments from the last paragraph. Therefore, for large datasets, we can apply the analytic estimation of the FAP for the periodogram $z_3(f)$ and other non-linear likelihood ratio periodograms in the same way as it was described in the previous sections for least-squares periodograms with known noise uncertainties.

6 PRACTICAL EXAMPLES OF NON-LINEAR PERIDOGRAMS

6.1 Detecting periodic signal with a fixed non-sinusoidal shape

Let us consider the case $d = 2$ with $\boldsymbol{\theta}$ incorporating only the amplitude and phase of the periodic signal to be detected. That is,

$$\mu = Kg(2\pi ft + \lambda), \quad (53)$$

where the 2π -periodic function $g(x)$ is given a priori and is centred so that $\bar{g} = 0$. This function determines the shape of the putative periodic variation.

Notice that the term M_{boundary} in (17) may now only appear due to the boundary points of the frequency segment. The phase λ is a periodic parameter defined over the self-closed circle which basically does not have a boundary. In other words, all maxima of η over λ are local maxima where $\partial\eta/\partial\lambda = 0$; no other maxima are possible in λ .

In this simple case we express FAP separately for the fixed-frequency and unknown-frequency cases, using the approach of Section 4.1. In the fixed frequency case, we find

$$\text{FAP}_{\text{single}} \lesssim M_{\text{single}}(z) \approx \sqrt{q}e^{-z}. \quad (54)$$

In the case of unknown frequency,

$$\text{FAP}_{\text{max}} \lesssim M_{\text{max}}(z) + M_{\text{single}}(z) \approx qWe^{-z}\sqrt{z} + \sqrt{q}e^{-z}. \quad (55)$$

The term M_{boundary} is equal to M_{single} here, since we have two end points of the segment $[0, f_{\text{max}}]$ and each should be counted as half.³ This term can be safely neglected in (55) anyway.

Notice that as long as the approximation of the uniform phase coverage is valid, the matrix \mathbf{G} appears here almost constant, so that we can use the simplified formulae (24). This means that although in the right-hand side of (55) we omitted a term of the order of e^{-z}/\sqrt{z} , corresponding to the term with a_1 of (20), the coefficient a_1 is itself negligibly small. As we have discussed in (Baluev 2008) for the sinusoidal model, the approximation of the uniform phase coverage works well even for time series with ultimately strong spectral leakage. This is because the aliasing/leakage-induced errors are concentrated only within a few narrow frequency segments, and they thus have only a negligible effect on the quantities expressed by an integral over a large frequency band (like A). However, for non-sinusoidal signals this approximation may appear poor due to reasons unrelated to the spectral leakage effects; namely it may be poor when $g(x)$ demonstrates narrow peaks that our observations are unable to cover with enough dense sampling.

In the general case $q \geq 1$. This inequality can be clearly derived by means of applying the Parseval identity to the Fourier series for $g(x)$ and $g'(x)$. When g is a harmonic function, we deal with the LS periodogram. In this case, $\bar{g}^2 = \bar{g}'^2 = 1/2$, and $q = 1$. This allows us to entirely reproduce the exponential single value distribution of the LS periodogram, $\text{FAP}_{\text{single}} = e^{-z}$, and the Davies bound $\text{FAP}_{\text{max}} \lesssim We^{-z}\sqrt{z} + e^{-z}$ from the paper (Baluev 2008). When $g(x)$ contains at least two Fourier terms we have $q > 1$, so the minimum $q = 1$ is attained for the sinusoidal and only sinusoidal variation.

Just as a non-trivial example, let us consider the case when $g(x)$ has a sawtooth shape: during the first half of its period it decreases linearly from 1 to -1 and during the second half it increases linearly from -1 to 1. In this case, $\bar{g}'^2 = 4/\pi^2$, $\bar{g}^2 = 1/3$, and $q = 12/\pi^2 \approx 1.216$.

6.2 The von Mises periodogram

Let us assume the following non-linear model for the periodic signal:

$$g(x, \nu) = \exp(\nu \cos x) - I_0(\nu), \quad \nu > 0. \quad (56)$$

In this definition I_0 stands for the modified Bessel function; we need it to satisfy the condition $\bar{g} = 0$. We can see that this function is 2π -periodic in x . At $\nu = 0$ we have a formal singularity because $g(x, 0) \equiv 0$ and (10) becomes degenerate. We can easily remove this degeneracy by making a replace

$$\tilde{g}(x, \nu) = \frac{g(x, \nu)}{\nu} = \cos x + \frac{\nu}{2} \left(\cos^2 x - \frac{1}{2} \right) + \mathcal{O}(\nu^2),$$

³ This is because there is a 50/50 chance that such a boundary value is actually a boundary minimum rather than a maximum, depending on the sign of the derivative in this end point.

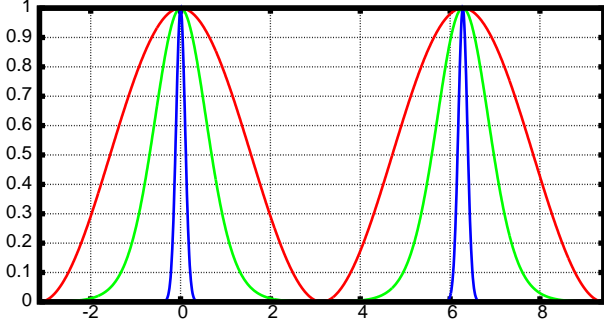


Figure 2. The graphs of the von Mises function $g(x, \nu)$ for $\nu = 0$ (the sinusoid), $\nu = 3$, and $\nu = 100$ (the most peaky case). All plots are prescaled to always cover the range $[0, 1]$ in the abscissa.

$$\tilde{K} = K\nu, \quad (57)$$

so that for $\nu \rightarrow 0$ our model becomes equivalent to a simple sinusoid. For large ν the function (56) represents a comb-like sequence of periodic narrow peaks, each having the width $\sim 1/\sqrt{\nu}$. Note that a very similar function with a bit different scaling, $\exp(\nu \cos x)/I_0(\nu)$, represents a probability density function of the so-called von Mises distribution. This distribution is a periodic analog of the Gaussian one, possessing a similar maximum-entropy property on a circle. For the sake of convinience we will also call (56) as the von Mises function, since these small differences in the centering and normalization are not very important for us.

In Fig. 2, we plot the von Mises function for several values of the localization parameter ν . Looking at these plots, we might notice that such shape may provide a satisfactory generic approximation to many physical variabilities that emerge in the astronomical practice. In particular, it may appear good for the lightcurves of many variable stars and planetary transits (just turn these graphs upside-down to obtain something similar to a transit). Since such model is functionally very simple (and thus quickly calculatable and easy in various analytic manipulations), it looks rather tempting to construct a periodogram that could utilise it as a model of the probe periodic signal.

Assume that we scan this periodogram in a rectangle $\nu_{\min} \leq \nu \leq \nu_{\max}$ and $0 \leq f \leq f_{\max}$, and disallowing the signal amplitude K to become negative. Then we should first evaluate the function τ using (20) for the interior of this rectangle (implying $n = 3$). It will be approximately proportional to W . Also, we should evaluate the function M_{boundary} , which contains the term responsible for the four sides and four vertices of the mentioned rectangle. Among these, only the terms due to the sides $\nu = \nu_{\min, \max}$ are important. This is because these are the only boundary terms proportional to W . The boundary terms due to the two other sides and due to the vertices do not contain this multiplier, because the frequency f is held fixed there. Since in practice W is large or very large, we only need to take into account the boundary edges running along the frequency axis.

The final approximation to the false alarm probability can be represented in the form

$$\text{FAP}(z) \lesssim M(z) = W e^{-z} \left[(X(\nu_{\max}) - X(\nu_{\min}))z + \right.$$

$$\left. + (Y(\nu_{\min}) + Y(\nu_{\max})) \frac{\sqrt{z}}{2} + \mathcal{O}(z^0) \right], \quad (58)$$

where

$$X(\nu) = \frac{1}{2W\pi^2} \int_0^{f_{\max}} df \int_0^\nu d\nu \int_0^{2\pi} d\lambda \sqrt{\det \mathbf{G}_{f\lambda\nu}(f, \lambda, \nu)},$$

$$Y(\nu) = \frac{1}{2W\pi^{3/2}} \int_0^{f_{\max}} df \int_0^{2\pi} d\lambda \sqrt{\det \mathbf{G}_{f\lambda}(f, \lambda, \nu)}. \quad (59)$$

In (58), the terms involving the factor X correspond to the local maxima of η in the interior of the rectangle. The difference $X(\nu_{\max}) - X(\nu_{\min})$ is because the factor A now contains an integral from ν_{\min} to ν_{\max} , while the function $X(\nu)$ is defined as an integral from 0 to ν . The terms with Y are for maxima on the two boundary lines $\nu = \nu_{\min}$ and $\nu = \nu_{\max}$. The sum $Y(\nu_{\min}) + Y(\nu_{\max})$ is because we need to sum the maxima at the both borders, and the extra multiplier of $1/2$ is because we should filter out half of these boundary maxima, due to the derivative $\partial\eta/\partial\nu$ having at these maxima an inappropriate sign with probability $1/2$. The quantities $WX(\nu)$ and $WY(\nu)$ represent, in fact, the factor A for the rectangle $[0, \nu] \times [0, f_{\max}]$ and for the boundary segments $\{\nu = \nu_{\min}, \nu_{\max}\} \times [0, f_{\max}]$.

In (59), the 3×3 matrix $\mathbf{G}_{f\lambda\nu}$ corresponds to the full gradient of η over all three non-linear parameters f, λ, ν ; the matrix $\mathbf{G}_{f\lambda}$ is a 2×2 submatrix of $\mathbf{G}_{f\lambda\nu}$ involving only the elements corresponding to only the parameters f and λ . Both these matrices depend on all three parameters. Notice that the terms in (58) containing Y basically correspond to the signal having a fixed non-sinusoidal shape (ν is fixed), and thus they can be also treated using the formalism of Sect. 6.1.

In the particular case $\nu_{\min} = 0$ (with the sign of K still fixed) we should take into account the obvious equality $X(0) = 0$:

$$\text{FAP}(z) \lesssim M(z) = W e^{-z} \left[zX(\nu_{\max}) + \right.$$

$$\left. + (Y(\nu_{\max}) + Y(0)) \frac{\sqrt{z}}{2} + \mathcal{O}(z^0) \right]. \quad (60)$$

Notice that in practice $Y(0) \approx 1$ with good precision (see below).

When K is allowed to be positive as well as negative we should double the right-hand side of the expression in (58). This is because we should now honour the local minima of the random field η too, as well as its local maxima. We have in this case:

$$\text{FAP}(z) \lesssim M(z) = W e^{-z} \left[2(X(\nu_{\max}) - X(\nu_{\min}))z + \right.$$

$$\left. + (Y(\nu_{\min}) + Y(\nu_{\max}))\sqrt{z} + \mathcal{O}(z^0) \right], \quad (61)$$

A special case occurs when $\nu_{\min} = 0$ and the sign of K is arbitrary. Then we have, basically, some degeneracy of the free variables at $\nu = 0$, making the cases $K < 0$ and $K > 0$ equivalent to each other (due to the symmetry of the sinusoid). This property can be used to refine the Rice bound a bit. Assume that we have some point at the boundary $\nu = 0$, such that $\eta > 0$ (implying $K > 0$) and $\partial\eta/\partial\nu < 0$. It is easy to show that at a dual point with $\lambda \mapsto \lambda + \pi$ the value of η changes the sign (hence $K < 0$), but the value of $\partial\eta/\partial\nu$ remains exactly the same. This is because the derivative of \tilde{g}

in (57) over ν is a π -periodic function of x (at $\nu = 0$), while the model \tilde{g} itself is 2π -periodic. Therefore, the derivative $\partial|\eta|/\partial\nu$ has different sign in these dual points, while the value of $|\eta|$ is identical.

Therefore, although there are many “good” boundary maxima at $\nu = 0$, which satisfy the condition $\partial|\eta|/\partial\nu < 0$ (meaning that $|\eta|$ necessarily decreases when we step from $\nu = 0$ inwards), each such a maximum has a dual “bad” maximum at $\lambda + \pi$, where $\partial|\eta|/\partial\nu > 0$. Since there are no constraints on λ , this means that for *any* boundary maximum at $\nu = 0$, either “good” or “bad”, we can find larger values of η in the interior $\nu > 0$. This implies that the *global* maximum of $|\eta|$ cannot be attained at the line $\nu = 0$. Following the terminology by Azaïs & Delmas (2002) (see their Theorem 2), we basically proved that the field ζ (or $|\eta|$) is a “field without boundary”, concerning the boundary line $\nu = 0$. This allows us to just drop the relevant boundary term with $Y(\nu_{\min} = 0)$:

$$\text{FAP}(z) \lesssim \tilde{M}(z) = W e^{-z} [2zX(\nu_{\max}) + Y(\nu_{\max})\sqrt{z} + \mathcal{O}(z^0)], \quad (62)$$

Notice that it is essential here that K has arbitrary sign, because otherwise we could not freely swap the values $\eta > 0$ with $\eta < 0$.

Let us first apply the approach of the Section 4.1 to evaluate X and Y . The derivatives of the function (56) look like

$$\begin{aligned} g'_x &= -\nu \exp(\nu \cos x) \sin x, \\ g'_\nu &= \exp(\nu \cos x) \cos x - I_1(\nu). \end{aligned} \quad (63)$$

Substituting them to (32), and using the well-known integral representations for the modified Bessel functions I_k , we obtain

$$\begin{aligned} \overline{g^2} &= I_0(2\nu) - I_0^2(\nu), \quad \overline{g_x'^2} = \frac{\nu^2}{2} [I_0(2\nu) - I_2(2\nu)], \\ \overline{g_\nu'^2} &= \frac{1}{2} [I_0(2\nu) + I_2(2\nu)] - I_1^2(\nu), \quad \overline{g_x' g_\nu'} = 0, \\ \overline{g g_\nu'} &= I_1(2\nu) - I_0(\nu)I_1(\nu), \end{aligned} \quad (64)$$

and then

$$\begin{aligned} q &= \frac{\nu^2}{2} \frac{I_0(2\nu) - I_2(2\nu)}{I_0(2\nu) - I_0^2(\nu)}, \quad v = 0, \\ \det \mathbf{R} = V &= \frac{I_0(2\nu) + I_2(2\nu) - 2I_1^2(\nu)}{2[I_0(2\nu) - I_0^2(\nu)]} - \\ &\quad - \left[\frac{I_1(2\nu) - I_0(\nu)I_1(\nu)}{I_0(2\nu) - I_0^2(\nu)} \right]^2. \end{aligned} \quad (65)$$

The behaviour of these quantities is not obvious from these formulae, so we need to understand their asymptotic behaviour. For small ν we can use the Taylor expansion of $I_k(z)$ to find that $q(0) = 1$ and $V(0) = 1/16$. This implies that for the factor X the integrand $q\sqrt{\det \mathbf{R}}$ is equal to $1/4$ at $\nu = 0$; for the factor Y we have $\det \mathbf{R} = 1$ by definition and $q\sqrt{\det \mathbf{R}}$ at $\nu = 0$ is unit. For large ν we may use the following asymptotically converging expansion:

$$I_k(z) \simeq \frac{e^z}{\sqrt{2\pi z}} \left(1 - \frac{4k^2 - 1}{8z} + \frac{(4k^2 - 1)(4k^2 - 9)}{2!(8z)^2} + \dots \right), \quad (66)$$

which can be found e.g. in (Korn & Korn 1968, §21.8). The calculations lead us to

$$q \simeq \frac{\nu}{2}, \quad V \simeq \frac{1}{8\nu^2},$$

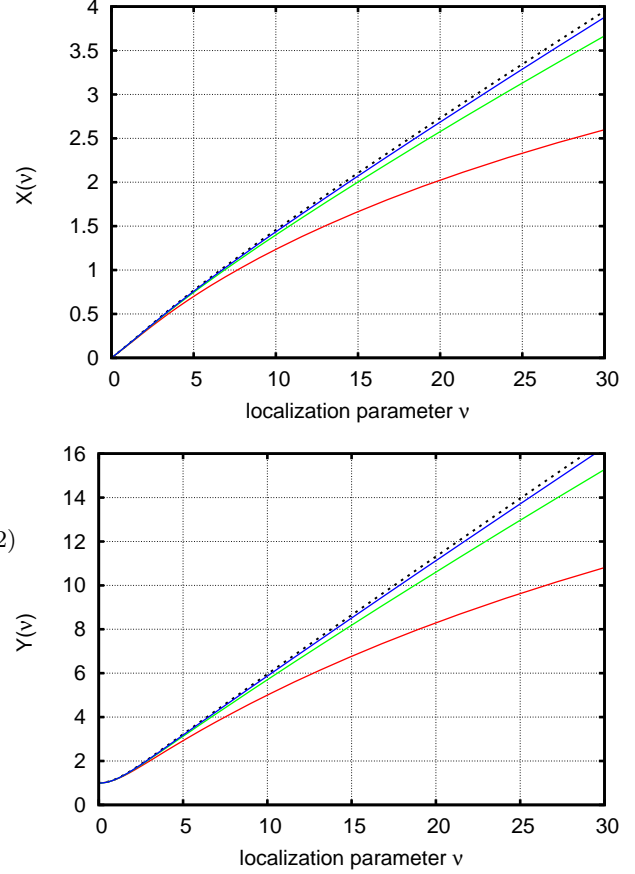


Figure 3. The coefficients X and Y for the von Mises periodogram. Dashed curve is for the approximation of Section 4.1, while the three solid curves correspond to the direct precise method of Section 4.2. These three curves (from down to up) correspond to three simulated time series, containing $N = 30$, 100, and 1000 randomly distributed simulated observations. For all cases we have $W \approx 100$.

$$q\sqrt{\det \mathbf{R}} \simeq \frac{1}{4\sqrt{2}} \quad (\text{for } X), \quad q\sqrt{\det \mathbf{R}} \simeq \frac{\nu}{2} \quad (\text{for } Y). \quad (67)$$

When calculating X , we should further integrate the relevant quantity over ν , so we obtain $X(\nu) \sim \nu$ for large ν . The factor Y does not request this integration, but its asymptotics appears eventually the same, $Y(\nu) \sim \nu$ for large ν . Also, for large ν we have $Y(\nu)/X(\nu) \simeq 2\sqrt{2\pi}$, hence the “primary” X -term in (58) really exceeds the Y -term only for $z > 2\pi$, and even beyond this level they remain mutually comparable up to rather large z . This means that both these terms should be taken into account in practice; none can be neglected.

These also results infer that the integral in (38) is infinite if we do not limit ν from the upper side. In this regard the parameter ν is similar to the frequency f . When we scan the usual periodogram in a wider frequency range, the probability to find a high noisy peak in this range inevitably increases. Similarly, an attempt to detect more quickly-varying signals with larger ν will inevitably increase our chances to catch a noisy fluctuation instead of a true signal.

We find that the approximate expressions for the factors X and Y obtained using the formalism of Section 4.1 a very accurate for small ν , but this accuracy decreases when ν

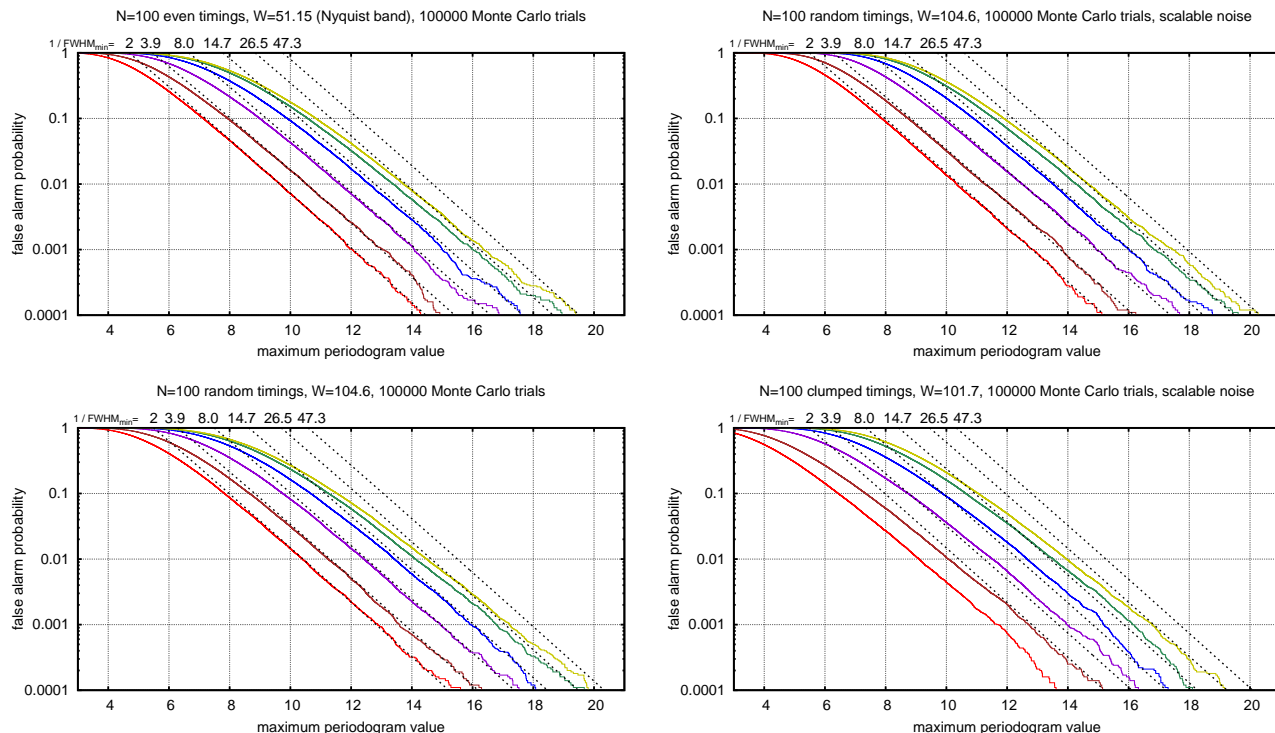


Figure 4. The FAP approximation of the von Mises periodogram in comparison with Monte Carlo simulations. In each panel we show the simulated and analytic FAP curves for $\nu_{\min} = 0$ and six different values of ν_{\max} from 0 to ≈ 315 . In the graphs we mark, instead of the upper limit for ν , a lower limit for a more intuitive FWHM (Full Width at Half Maximum) characteristic of the signal peaks. It can be easily mapped one-to-one with ν_{\max} and it varies here from $1/2$ (corresponding to the sinusoidal variation) to approximately $1/50$ (typical for e.g. a planetary transit). The panels to the left correspond to the cases with the noise uncertainties σ_i known a priori; the ones to the right are for the multiplicative noise model of Sect. 5, $\sigma_i^2 \propto 1/w_i$. In the case of “clumped random timings” (right-bottom panel) the $N = 100$ points of the time series were equally split in 10 equidistant groups with 90% gaps between them. This implies a very strong aliasing, which makes our analytic approximation for FAP relatively inaccurate, though they still work as an upper limit.

grows, and increases when N grows (Fig. 3). We assume that the error of this approximation emerges because our N observations cannot sample well the narrow peaks of the signal having the width $\sim 1/\sqrt{\nu}$.

We have done some Monte Carlo simulations to test the quality of the approximation (60). The results are shown in Fig. 4. We find that our analytic approximations behave exactly as we might expect. They have good accuracy for practically important small levels of the FAP and for not too large ν_{\max} and time-series leakage. For larger ν_{\max} or for strong leakage the accuracy somewhat degrades, but the formula (60) still works as an upper bound on FAP. Our conclusion is that this formula would be certainly useful in practical applications.

At last we would like to demonstrate the power of the von Mises periodogram itself. We generated a simulated time series with $N = 1000$ randomly spaced observations. The values of the simulated measurements contained two periodic signals with comparable amplitudes: a sinusoidal variation and periodic flat drops (simulating planetary transits). Both signals were below the noise level, so the noise should provide significant contamination. The von Mises periodogram of these data is plotted in Fig. 5. We can see that it allows an easy detection of the both signals at once ($f \approx 8$ and $f \approx 56$), while the LS periodogram (which represents, basically the middle horizontal slice of the plot) would robustly

reveal only the sinusoidal periodicity, allowing the planetary transit to slip away until the next step of the analysis.

7 CONCLUSIONS

In this paper, we extended our previous results (Baluev 2008, 2009b) to the case when the model of the signal to be detected in the noisy data depends on unknown parameters in a non-linear manner. The definition of the periodogram was extended to this non-linear (and non-sinusoidal) case in terms of the χ^2 and likelihood-ratio tests. We described a generic method of constructing an asymptotic approximation to the false alarm probability. Based on these general results, we considered two specialized non-linear periodograms. The first one involves a fixed-shape periodic non-sinusoidal model of the signal. The second one models the signal with the so-called von Mises function $\exp(\nu \cos x)$. This function is very remarkable, because it allows fairly good approximation of very different periodic variations, from the plain sinusoid to a model with periodic narrow peaks or drops (typical for e.g. the exoplanetary transit lightcurve). For both these periodograms we provide a complete theoretical solution of the false alarm probability problem.

Moreover, for the von Mises periodogram we offer a supporting package of C++ programs, that may dramatically

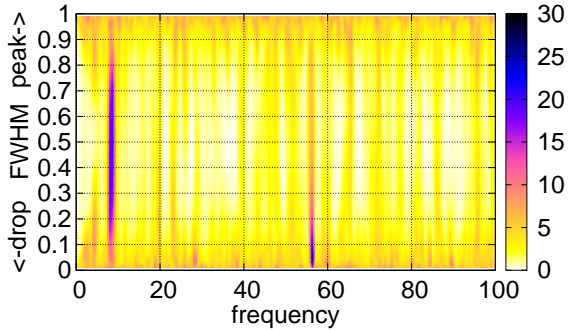


Figure 5. The von Mises periodogram plotted for a test time series containing two periodic signals contaminated by noise (see text). The lower half ($\text{FWHM} < 0.5$) of the graph correspond to a drop-type periodicity, while the upper part (for which we made a symbolic replacement $\text{FWHM} \mapsto 1 - \text{FWHM}$ for convenience, so the labelled values FWHM exceed 0.5) is for a peak-type variation. The slice at $\text{FWHM} = 0.5$ represents the LS periodogram assuming the sinusoidal model of the signal.

facilitate the use of the relevant theory in practice. This package is attached to the present paper as the online-only supporting material (as a compressed archive).

We expect that the results of this work can be used in a wide variety of astronomical applications that deal with non-sinusoidal periodicities in observational data. These research fields are ranged from the studies of variable stars to the studies of extrasolar planetary systems.

In the forthcoming and future work, we plan to apply our approach to the so-called double-frequency periodogram, where the signal is modelled by a sum of two independent sinusoidal terms (to appear in *Astron. Lett.*, under review), to the Schuster periodogram and to the so-called Keplerian periodogram introduced by Cumming (2004), also known as “2DKLS periodogram” (O’Toole et al. 2009).

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project 12-02-31119 mol.a) and by the Programme of the Presidium of Russian Academy of Sciences “Non-stationary phenomena in the objects of the Universe”. I would like to express my gratitude to the anonymous reviewer, who provided suggestions of a great value.

REFERENCES

- Azaïs J.-M., Delmas C., 2002, *Extremes*, 5, 181
 Azaïs J.-M., Wschebor M., 2009, *Level Sets and Extrema of Random Processes and Fields*. Wiley
 Baluev R. V., 2008, *MNRAS*, 385, 1279
 Baluev R. V., 2009a, *MNRAS*, 393, 969
 Baluev R. V., 2009b, *MNRAS*, 395, 1541
 Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, 304, 15

- Cumming A., 2004, *MNRAS*, 354, 1165
 Cumming A., Marcy G. W., Butler R. P., 1999, *ApJ*, 526, 890
 Dacunha-Castelle D., Gassiat E., 1999, *Ann. Stat.*, 27, 1178
 Ferraz-Mello S., 1981, *AJ*, 86, 619
 Frescura F. A. M., Engelbrecht C. A., Frank B. S., 2008, *MNRAS*, 388, 1693
 Horne J. H., Baliunas S. L., 1986, *ApJ*, 302, 757
 Koen C., 1990, *ApJ*, 348, 700
 Korn G. A., Korn T. M., 1968, *Mathematical Handbook for Scientists and Engineers*. McGraw Hill, New York
 Kratz M. F., 2006, *Probability Surveys*, 3, 230
 Lomb N. R., 1976, *Ap&SS*, 39, 447
 O’Toole S. J., Jones H. R. A., Tinney C. G., Butler R. P., Marcy G. W., Carter B., Bailey J., Wittenmyer R. A., 2009, *ApJ*, 701, 1732
 Rice S. O., 1944, *Bell System Tech. J.*, 23, 282
 Scargle J. D., 1982, *ApJ*, 263, 835
 Schwarzenberg-Czerny A., 1996, *ApJ*, 460, L107
 Schwarzenberg-Czerny A., 1998a, *MNRAS*, 301, 831
 Schwarzenberg-Czerny A., 1998b, *Baltic Astron.*, 7, 43
 Zechmeister M., Kürster M., 2009, *A&A*, 496, 577

This paper has been typeset from a \TeX / \LaTeX file prepared by the author.